

A Mixture Model of Global Internet Capacity Distributions

Hyunjin Seo

Assistant Professor and Docking Faculty Scholar
William Allen White School of Journalism and Mass
Communications
University of Kansas
1435 Jayhawk Blvd.,
Lawrence, KS 66045-7515
E-mail: hseo@ku.edu

Stuart Thorson

Donald P. and Margaret Curry Gregg Professor
The Maxwell School
Syracuse University
409 Crouse-Hinds Hall
Syracuse, NY 13244-1075
E-mail: thorson@syr.edu

Accepted for publication in *Journal of the Association for Information Science and Technology*

Corresponding author: Hyunjin Seo, hseo@ku.edu

This paper develops a preferential attachment based mixture model of global Internet bandwidth and investigates it in the context of observed bandwidth distributions between 2002 and 2011. Our longitudinal analysis shows, among other things, that the bandwidth share distributions – and thus bandwidth differences – exhibit considerable path dependence where country proportions of international bandwidth in 2011 can be substantially accounted for by a preferential attachment based mixture of micro-level processes. Interestingly, our preferential attachment model, consistent with empirical data, does *not* predict increasing concentration of bandwidth within top ranked countries. We argue that recognizing the strong, but nuanced, historical inertia of bandwidth distributions is helpful in better discriminating among competing theoretical perspectives on the global digital divide as well as in clarifying policy discussions related to gaps between bandwidth rich and bandwidth poor countries.

Keywords: Internet, international bandwidth distribution, digital divide, preferential attachment, mixture model, path dependence

Introduction

The development of the global Internet as a medium for high-speed global communications has had significant consequences for aspects of society including collaborations within and across borders, political movements, social changes, and trade (Benkler, 2006; Castells, 2011). The Internet plays a crucial role in diffusion of innovations and knowledge (Bonfadelli, 2002; Hassani, 2006; Ioannides, 2013; Jovanovic & Rob, 1989), political and civic engagement (Howard & Hussain, 2011; Norris, 2003), and economic development (Heeks & Kenny, 2001; Manyika & Roxburgh, 2011). Social interactions, which have significant implications for multiple facets of society at local, national, and international levels, are often facilitated by the Internet and online social networking sites (Benkler, 2006; Ioannides, 2013; Zenou & Jackson, 2013). Therefore, the strength of a country's international Internet connections with other countries can be expected to have significant consequences for its future development.

These connections are not uniformly distributed across the globe. The so-called global digital divide – differences between countries in terms of access to and use of information and communication technologies (ICTs) – has been the sub-

ject of active empirical investigation (Asia Pacific Economic Cooperation, 2002; Chinn & Fairlie, 2010; Guillén & Suárez, 2005; Hilbert & López, 2011; International Telecommunications Union, 2012; Mossberger, Tolbert, & Hamilton, 2012; Pejovic et al., 2012; Van Dijk, 2005; Wunsch-Vincent & Vickery, 2010). These studies have found significant differences in Internet access and/or use between countries. Policy questions remain as to whether differences are widening or narrowing, what factors cause the differences, and even how best to conceptualize these differences (James, 2007; Yu, 2006). Many studies have considered economic, sociological, political, and regulatory characteristics of countries to understand cross-national differences in Internet assets (Andrés, Cuberes, Diouf, & Serebrisky, 2010; Asia Pacific Economic Cooperation, 2002; Chinn & Fairlie, 2010; Guillén & Suárez, 2005; Roycroft & Anantho, 2003; Van Dijk, 2005; Wunsch-Vincent & Vickery, 2010). However, little research has focused on how distributions of global Internet capacity have evolved and how changes in the distributions might help us understand growth patterns of the global Internet. We highlight this, as the notion of digital divide is inherently a distributional concept. To account for national differences in Internet assets we must understand both the global distribu-

tion of those assets and how, if at all, characteristics of these distributions change over time.

We propose a preferential attachment based mixture model of country to country Internet bandwidth distributions and examine it in the context of observed bandwidth distributions from 2002 to 2011. The dataset analyzed is built from bandwidth capacity data for direct country-to-country Internet connections in each year for the period 2002 to 2011. Bandwidth capacity, the contracted amount of capacity for transfer of Internet data, serves as an upper bound on the total amount of information that can be moved internationally over the Internet.

We are interested in the recent evolution of the global communications network as measured by the distribution of international bandwidth capacity. This approach permits empirical investigation of the impact of initial bandwidth distribution on subsequent development. Empirically tested estimates of specific forms of growth are essential to discriminating competing theoretical perspectives as well as for clarifying policy discussions related to different notions of how, if at all, to address differences between bandwidth rich and bandwidth poor countries. In addition, our model development offers a methodology for helping advance the understanding and visualization of global communication networks and distributional differences.

Literature Review

Global Internet Capacity

Previous research showed that global Internet connectedness has grown significantly over the past decade with the global Internet network becoming denser (*1 citation removed for masked review*). Growth has not been equally distributed and these inequalities have implications for political, economic, and social developments across countries, regions, and continents. For this reason, the global digital divide, gaps between countries in terms of access to and use of ICTs, has attracted wide attention from scholars and policy makers (Chinn & Fairlie, 2006; Guillén & Suárez, 2005; Norris, 2003; Van Dijk, 2005). Previous studies have investigated how Internet assets have spread and developed at different rates and levels across the globe. In particular, they examined economic, political, social, policy, and/or technological variables to understand what factors influence different levels of technology adoption and deployment in specific countries.

Infrastructure and economic advancement have generally been identified as some of the most important factors related to the global digital divide (Asia Pacific Economic Cooperation, 2002; Chinn & Fairlie, 2006; Kiiski & Pohjola, 2002; Norris, 2003). For example, Chinn and Fairlie (2006) found that per capita income, urbanization rate, and telecommunications infrastructure were significantly associated with personal computer and Internet penetration. Kiiski and Pohjola

(2002) examined 60 OECD and developing countries and found per capita income and prices of telephone access to be statistically significant correlates of Internet hosts per 1,000 inhabitants. A study of Asia Pacific Economic Cooperation (APEC) member countries identified per capita income and urban population proportion as significant determinants of the number of Internet users per capita (Asia Pacific Economic Cooperation, 2002).

In addition, country specific political and institutional environments are associated with access to and usage of ICTs (Asia Pacific Economic Cooperation, 2002; Dasgupta, Lall, & Wheeler, 2005; C. Lee & Chan-Olmsted, 2004). APEC (Asia Pacific Economic Cooperation, 2002) found that political freedom was positively correlated with Internet penetration in 21 APEC member countries. Scholars have also noted that political institutions play a key role in shaping adoption, diffusion, and use of the Internet (Dasgupta et al., 2005). Level of education is another factor positively associated with access to and use of Internet assets in different countries (Asia Pacific Economic Cooperation, 2002; Chinn & Fairlie, 2006; Kiiski & Pohjola, 2002; Norris, 2003; Quibria, Ahmed, Tschang, & Reyes-Macasaquit, 2003). Based on a cross-national analysis of Asian countries' use of computer, the Internet, telephone, and mobile phones, Quibria et al. (2003) found that education, along with income and infrastructure, played a significant role in shaping the digital divide.

While there is a general consensus that striking disparities exist between countries in access to and use of Internet assets, there remain disagreements on the evolution of those disparities. Guillén and Suárez (2005, p. 697) argue that significant differences between countries will remain because "they are the result of the fundamental economic, political and social gap that separates the advanced from the less developed countries." They predict that countries with an initial advantage would benefit disproportionately from the Internet resulting in a rich-getting-richer phenomenon. Based on a comparison of Asian and non-Asian countries, Wong (2002) suggested that differences will remain wide and are perpetual, though the research also found the average gaps in ICT diffusion intensity between the Asian and non-Asian countries as well as among Asian countries to have narrowed slightly during the period of 1994-1998.

Fink and Kenny (2003) present a more optimistic view arguing that gaps are narrowing in that when a per-income measure of access to a variety of ICTs is used, developing countries "digitally leapfrog" developed countries and will ultimately catch up to the developed. They emphasize that relative rates of growth should be examined in determining whether differences between developed and developing countries are widening or narrowing. Analyzing the process of Internet diffusion in 214 countries between 1990 and 2004, Andrés et al. (2010) also concluded that low-income

countries are enjoying a faster growth rate than high-income countries despite an initial lag and that they are very slowly catching up.

Hilbert and López (2011) illustrate both the importance and difficulties in harmonizing empirical measures in this domain. Decisions to focus on particular measurements often depend on availability of data but sometimes also on political views of researchers (Yu, 2006). These differences may also result from the fact that studies in this area have almost exclusively focused on examining attributes of individual countries with less attention being paid to the possible role of both history and network relations in amplifying or reducing differences. We aim to address this by looking at differences between dyads of countries as well as properties of the overall distribution of global Internet bandwidth.

Growth Patterns of Communication Networks

Communication networks have received increased attention from both scholars and practitioners with rapid developments of ICTs offering actors at different levels of society more affordable and available ways to connect and interact. Communication networks refer to “patterns of contact that are created by the flow of messages among communicators through time and space” (Monge & Contractor, 2003, p. 3). In this increasingly interconnected society, most of our daily communication activities are embedded in Internet-based networks – whether keeping in touch with friends through Facebook, exchanging emails with your business partners, or searching information on the Web. Moreover, cross-national social relations have been taken to a new level thanks to information technologies (Benkler, 2006; Castells, 2011). People in disparate parts of the world work together to produce content on wikis – collaborative websites – and can share their opinion with widely distributed Internet users through social media such as blogs, Facebook, YouTube, and Twitter. There is empirical evidence that the Internet has facilitated communication between activists in different Arab Spring countries (*1 citation removed for masked review*). These are only recent examples showing how communication behavior is embedded in communication networks. The development in information technologies has been an important driving force of globalization which Giddens (1990, p. 64) defined as “the process of strengthening the worldwide social relations which link distant localities in such a way that local events are shaped by circumstances at other places in the world.” Studying the patterns of global Internet connectedness will help us better anticipate future communication behavior.

Communication scholars have investigated communication networks from the standpoint of the global Internet and international telecommunication (Barnett & Park, 2014; Barnett, Ruiz, Hammond, & Xin, 2013; Howard, 2010; S. Lee, Monge, Bar, & Matei, 2007; Park, Barnett, & Chung, 2011). For example, Barnett (2001) conducted a longitudinal anal-

ysis of the international telecommunications network from 1978 to 1996. His analysis showed that the international telecommunications network has become denser, more centralized, and highly integrated over time. A more updated picture of international telecommunications networks was offered by S. Lee et al. (2007). They found that the international telecommunication network has evolved toward a decentralized structure with increased clusters within the network and increasing connectivity within peripheral nations.

TeleGeography’s annual surveys of Internet traffic and capacity indicates that total raw international Internet bandwidth has increased from less than 1 Terabit per second (Tbps) in 2002 to about 55 Tbps in 2011 (TeleGeography, 2012). The economic impact of the Internet has been enormous with Manyika and Roxburgh (2011) estimating that it now accounts for an average of 3.4 percent of GDP for the world’s largest economies. Governments, either directly or indirectly, often provide funding for much of the Internet’s development (Hafner & Lyon, 1999; Howard, 2010; Standage, 1998). Barnett and Park (2005) have shown that the early pattern of international Internet connections largely mirrored political-economic structure at the time. That is, rich and powerful countries enjoyed high bandwidth direct connections with one another and with some less well off counties. In comparison, less powerful countries had lower bandwidth connections with geographic neighbors and, in some cases, one of the bandwidth rich countries.

How networks grow has been a key question, and by now it has been well documented that many interesting networks grow in a non-random manner with new assets in the network being distributed in proportion to their distribution at a previous time (Barabási, 2003; Newman, Barabási, & Watts, 2006). This dynamic can lead to a “rich getting richer” phenomenon. Preferential attachment is a mechanism often posited to account for this type of evolution. Under preferential attachment, the growth of a network is modeled as a stochastic process in which certain attributes are distributed to each node in a manner roughly probabilistically proportionate to the proportion of those of attributes that node had at the previous time with new nodes sometimes being added as the network evolves. Under appropriate assumption, network growth generated by preferential attachment results in right skewed heavy tailed distributions with a small percentage of nodes having a disproportionate share of the attribute being studied. In the context of the city size distribution, Eeckhout (2004) found that Pareto and lognormal differ most dramatically in the left part of the distribution. If we look at all the websites on the World Wide Web, we would find that most have relatively few links to other sites. A random sample of websites would contain mostly (if not only) such sites. However, a small proportion of websites such as google.com or naver.com will be linked to by a very large number of other,

generally smaller, sites. These large sites would be in the heavy right tail of the distribution of all websites.

Scholars have observed that many naturally occurring communication networks exhibit a heavy tailed distribution (Barabási, 2003; Hyun, 2012). In heavy tail discussions we look at the lower right part of the probability density function plot and refer to that as the tail. For example, Hyun (2012) found a heavy tailed distribution of incoming links for political blogs in the United States, Great Britain, and Germany.

While insufficient attention has been paid to analyzing how distributions of global Internet capacity have evolved and how changes in the distributions might help us understand growth patterns of the global Internet, several previous studies made theoretical or methodological advances related to this topic. Most recently, Hilbert (2014) analyzed a comprehensive database – based on more than 1,100 different data sources from international organizations – in examining inequalities in access to, usage of, and impact of digital information flows between 1986 and 2010. He found that the level of informational inequality began to decrease only recently (2006–2010) suggesting that “far-reaching social, economic, cultural, and political transformations of the information revolution are yet to come and require research attention” (p. 832). In analyzing the network structure of the international Internet, Barnett and Park (2014) used four different sources of data: (i) bilateral bandwidth between countries, (ii) hyperlink connections among nations’ domain names, (iii) structural equivalence of nations from the perspective of websites, and (iv) structural equivalence of nations from the national perspective. They found significant correlations between the hyperlinks, bandwidth and structural equivalence from the website concluding that they should be considered as “multiple indicators of the same network” (p. 573). Their findings also suggest that the international Internet network “appears to consist of series of small worlds determined by language, geography and historical circumstances” (p. 563).

This study continues in this tradition and analyzes patterns of international Internet bandwidth and connections to understand hierarchy and inequality in Internet connections across the globe. Specifically, we examine how growth patterns of the global Internet have evolved during the period of 2002–2011 and propose a mechanism that may have generated those growth patterns.

Methods

We focus on the distribution of country bandwidth proportions over time. Insofar as discussions of the global digital divide are concerned, proportions are generally what are of concern as they provide a standard for comparing across both countries and years. If the bandwidth distribution were to condense upward over time it would be reflected in changing proportions with bandwidth richer countries gaining a higher proportion of the total bandwidth. Bandwidth amounts, on

the other hand, do not offer clear ways of over time comparisons. For example, changes in data compression algorithms and patterns of demand for types of digital data render cross time comparisons of absolute levels problematic. Unlike with macro economic statistics such as GNI, there are no generally accepted methods for adjusting annual bandwidth amounts to constant values. Thus we consider proportion bandwidth shares in current bandwidth levels.

The literature dealing with international Internet bandwidth has consistently found that politically powerful and economically rich countries have more bandwidth (amount and proportion) than do less well off countries (Barnett, 2001; Barnett & Park, 2014). This has meant that, as with many wealth-related distributions, bandwidth distributions have heavy right tails; most countries have relatively little bandwidth and a few have quite a lot.

Our initial theoretical approach to accounting for bandwidth distributions assumes that as new international bandwidth is created the likelihood that it will accrue to a particular country will, all other things being equal, be proportional to the amount that country had at a previous time. Depending upon precisely how these assignments are done the results can result in a rich-getting-richer dynamic in which there is increased condensation of bandwidth proportions. More generally such a preferential attachment mechanism suggests that current distributions are derived from past ones and thus exhibit path dependence (Arrow, 2000; Arthur, 1994; Pierson, 2000).

Vinciguerra, Frenken, and Valente (2010) tested a preferential attachment model against empirical data on the European Internet infrastructure network in examining whether geographic distance and country borders can explain the structure of the network. Their analysis found that “the preferential attachment logic of network growth, combined with barriers to connect due to geographical distance and country borders, is successfully reproducing a significant part of an infrastructure network” (p. 1978). They point out that “the effect of specific historical events” is an important topic that has not been explored in their research or other previous studies and emphasized the importance of understanding “whether the entry of particular cities early in the network has had a long-lasting effect on the resulting network structure that emerged later on” (p. 1981).

To examine preferential attachment with regard to international Internet bandwidth we specify a simple urn-based mechanism credited to Pólya (Chung, Handjani, & Jungreis, 2003; Eggenberger & Pólya, 1923).¹ The basic urn problem begins with a finite number of urns each containing one ball. At each new time point, with some probability q , a new urn is created and a new ball is placed in that urn. Or, with some probability $p01 - q$, the new ball is placed in an existing urn

¹Our notation is based upon that proposed in Chung et al. (2003).

so that the probability that it is placed in a given urn is proportionate to m^γ where m is the number of balls already in that urn.

We reinterpret the urn model in terms of countries and quanta of bandwidth. Each urn becomes a country and each ball a quantum of international bandwidth. The global distribution of bandwidth quanta at a particular time would then correspond to the number of balls in the various urns. We consider the special case where $q = 0$. That is, new quanta of bandwidth are allocated only among existing countries². In addition, rather than starting the process with only one marble in each urn we will assume an initial unequal distribution of marbles. If we stipulate that 1Mbps (Megabit per second) is the size of a single bandwidth quantum, then in 2002 the countries ranged from holding 1 quantum to containing 369,360 quanta.

The m^γ component of the model captures *preferential attachment* and we focus on the simplest case where $\gamma = 1$. In other words, when a new quantum of bandwidth is allocated the *probability* that it will go to a given country is directly proportional to the share of total bandwidth quanta the country currently possesses. Taken together, these assumptions produce what is called a *finite Pólya Process*.

The particular allocation path taken is itself the result of a sequence of probabilistic outcomes and any particular path is but one instance of how allocations might have occurred. There is a distribution of paths which might occur given the model and the initial conditions. The question of system evolution can be investigated by identifying the probability density function, PDF, associated with the various paths under given initial conditions. Each run of the system generates a PDF. Suppose there are n countries. Each country will have some non-zero percentage of the total bandwidth quanta. These percentages can, after dividing by 100, also be interpreted as a length n vector of probabilities whose sum is 1.0. That vector contains the country shares of the empirical PDF for that point in the evolution of the process. Given the stochastic nature of the mechanism, each run could produce different results. What we are really interested in is, over the long run, characteristics of the probability distribution of these vectors; a PDF over PDFs.

Fortunately, our urn mechanism has several useful properties. First, the vector of balls in urns generated by this model forms a homogeneous Markov chain. It is Markovian since at each time step the urn chosen at random depends only on the distribution within the urns at the previous time. Homogeneity follows from the probability transition matrix not depending upon a time index. That the distribution vector is a Markov chain formalizes the path dependence characteristic noted above.

In our case, the vector of balls corresponds to the vector of (modeled) bandwidth quanta. However, our interest is with the proportion of bandwidth held by each country. This cor-

responds to the proportion of the total number of balls each urn has at a given point in time. The vector of proportions is not a Markov chain since the ratio value at a time point will depend not on the ratios at the previous time but rather on the actual number of balls in each urn.

It has been shown that the the sequence of ratios is a Martingale (Mahmoud, 2008) and this has been used to prove that as the mechanism evolves, the expected ratios will stabilize. Even more usefully, they will follow the Dirichlet distribution whose parameters are the initial number of balls in each urn.

In our case where there are 184 countries, the distribution of PDFs is described by a Dirichlet distribution, *Dir*, with parameters $\alpha_1, \alpha_2, \dots, \alpha_{184}$ where α_i denotes the initial bandwidth of the i th country. The distribution of the individual PDFs depends upon the vector of initial α values.

One implication of the model is that while the distribution of bandwidth proportions will move around (how much is characterized by the underlying Dirichlet distribution of which it is a realization) it should, in an expected value sense, ultimately stabilize around the initial distribution of bandwidth. This may be counterintuitive given that preferential attachment mechanisms often generate distributions with increasing concentration over time. In our case the result follows from the not unrealistic assumption that new urns (countries) are not created. In model terms, $q = 0$. If $q > 0$, then new countries could be created to receive new bandwidth. An implication is that these late-comers would initially have a very small amount of bandwidth. Additionally, they would be expected to continue to have small proportion of the total as late-comers tend to connect to higher bandwidth nodes but very few new nodes attach to more recent low bandwidth ones. Over time this would have the consequence of increasing the concentration of bandwidth. Such a mechanism can be shown to generate power-law distributions. Indeed, following Piketty (2014) one might speculate that some portion of the growing personal wealth inequalities may be attributable to population changes.

The macroscopic bandwidth amounts we observe are the result of a number of unobserved micro-processes. These might include preferential attachment (as above) economics, politics, private sector decisions, and so on. As Frank (2014, p. 3) notes, in such cases, “The greater the number of small scale fluctuations that combine to form an aggregate, the greater the total randomness in the macroscopic system.” Nor are these micro-processes contained within a single country. The bandwidth between, say, South Korea and

²This is similar, though not identical to the situation in the global Internet where there were 186 countries in 2002 and 201 countries by 2002. In our empirical analysis we considered only the 184 countries with reported bandwidth for the entire 10 year period of study so we basically, albeit somewhat artificially, meet this condition.

Japan will depend upon processes in both countries as well as possibly similar processes in other countries connecting to Japan and/or Korea through Japan.

The distribution of bandwidth observed in the TeleGeography data results from the aggregation of these micro-processes. Characteristics of the probability distributions of the observed macroscopic events follow from the manner in which the sub-processes are aggregated. For example, the micro-processes might be multiplicative in which case we might expect a lognormal distribution at the macroscopic level. Or more simply, the micro-level processes may be additive. In such a case, the mean effect of each process can be summed to give a maximum likelihood estimate of the net result for each country of all those processes. In what follows, we will assume additivity.

Under this assumption, we can expect the outcomes will have a Gaussian distribution with the micro-level mean as the MLE for each country. Suppose, however, that within our 184 countries there are some distinct categories of micro-processes that, while Gaussian, have different parameters (mean, variance, and relative weight). Then the overall process we are observing at the macroscopic level (our data) would result from a weighted mixture of those normal distributions.

The model we will examine has at its core the urn approach to account for preferential attachment but will be supplemented by the possibility that some countries will be able to make sustained efforts to increase (or lose) their share beyond that predicted by simple preferential attachment. This could be thought of as arguing that a country's bandwidth will be the sum of the urn prediction plus the (possibly negative) result of specific policy, human resource, and economic factors. And, of course, since the sum of probabilities must sum to 1.0 a country's bandwidth share may be decremented by a normalization factor.

If the expected mean effect of all the subprocesses were equal to zero for all countries, then we simply have preferential attachment as in the basic urn model. Alternatively, if each country's bandwidth (proportion) were almost completely the result of unique policy effects, then preferential attachment would play a decreasingly important role over time. However, if a small, relative to the number of countries, distinct patterns of policy effect can be identified then this could produce something similar to convergence clubs of the sort discussed by Galor (1996).

Formally, let f_{2011} be the overall bandwidth distribution for 2011, θ_k be the mean and standard deviation for the for the k^{th} (in our case) normal mixture distribution and λ_k be the weight or contribution of the k^{th} distribution to the overall bandwidth distribution we have:

$$f_{2011}(x) = \sum_{k=1}^K \lambda_k f(x; \theta_k) \quad (1)$$

The result is a basic urn model which would predict that, for example, 2011 share distributions would look very much like those in 2002 absent sustained policy interventions. And, to the extent there were deviations, we would investigate whether a parsimonious grouping of countries, based upon a Gaussian mixture of the form in Equation 1, sheds substantive light on the deviations. A diagram of the model is in the Appendix.

Data

The data used in this research consists of country level international bandwidth data as curated by TeleGeography (2012) and covers the period beginning with 2002 and ending with 2011. International Internet bandwidth refers to the amount of data that can be transferred over the Internet across national borders in a given amount of time, and thus, for each direct connection, is an upper bound on direct usable country to country Internet traffic flow. Bandwidth amounts reflect the total capacity of the connections each country has with one another and in that sense are an indicator of a country's position in the global Internet. This focus on bandwidth amounts could be misleading if countries with larger bandwidth simply had huge connections with one or two other countries. However, such is not the case. Correlations between country degree ranking and country bandwidth during the period of 2002-2011 show high ranking bandwidth countries tended to also be high ranking degree countries.

Results

The first year or which we could find country level data for what corresponds to the full set of countries in today's Internet was 2002. The number of countries with a measured direct connection to at least one other country was 186 in that year and grew to 202 by 2011. We consider the 184 countries which had connections in each of the years. This should not have much of an effect on our analysis as the excluded countries had very little bandwidth in the years for which they had connections. The total bandwidth of the 184 countries in any of the years was never less than 99.9 percent of the total reported international bandwidth in that year.

During this period, total international Internet bandwidth grew from slightly less than one terabit per second (1Tbps) to almost 55Tbps by 2011. Moreover, the growth has been roughly exponential over the period as shown in Figure 1 where bandwidth, plotted on a logarithmic Y axis, appears linear. The heavy line represents the best (linear) fit given the data.

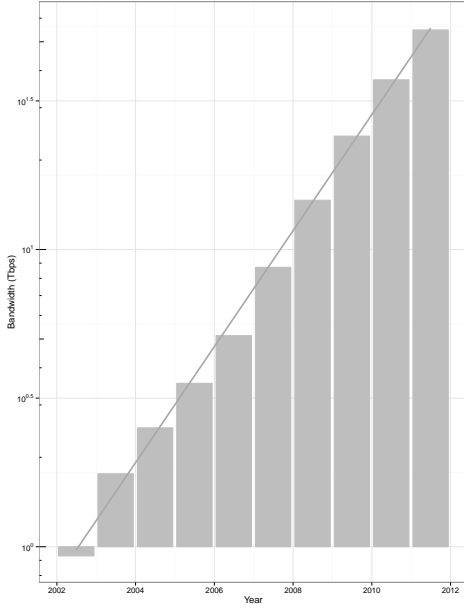


Figure 1. Total International Internet Bandwidth 2002-2011

Note: Y axis is on a logarithmic scale.

In evaluating how well our model accounts for the available data we will consider a rather hard test in that we examine the degree to which our model using 2002 data can account for 2011 bandwidth shares. In the simplest case we ignore the policy variable and assume that by 2002 the share probabilities had stabilized.³

Figure 2 is a scatter plot of the 2002-2011 country shares against the 2011 shares along with the least squares line. While the overall pattern and slope is in the predicted direction, the points do not all fall on the line as would be predicted by the simple urn model. With this volume of bandwidth quanta the differences are well outside those that could be expected to occur by chance alone. Moreover, visual inspection of the plots suggests that the smaller the lag, the more closely the lagged distribution approximates the 2011 distribution. Yet the simple urn model would suggest that after a large amount of bandwidth had been distributed, the shares should stabilize and the number of lag years should not make much difference. The largest residuals come from countries such as Germany, China, and Russia. There are specific historical, economic, or policy factors that may have influenced changes in the bandwidth shares of these countries. A more detailed discussion on this is provided later in this section in the context of the urn-based mixture model.

Moreover, Figure 2 illustrates the skew of the share distribution with most countries clustered at the low end for all of the years. One possibility is that, contrary to the simple urn model, there is in fact tail behavior consistent with a power law or a lognormal distribution. We tested for this and reported results in (1 citation removed for masked review).

Using the Clauset, Shalizi, and Newman (2009) method we found little compelling evidence in support of either a power law tail or lognormal distribution. In part the small number of cases in the tail made it difficult to reject much of anything.⁴

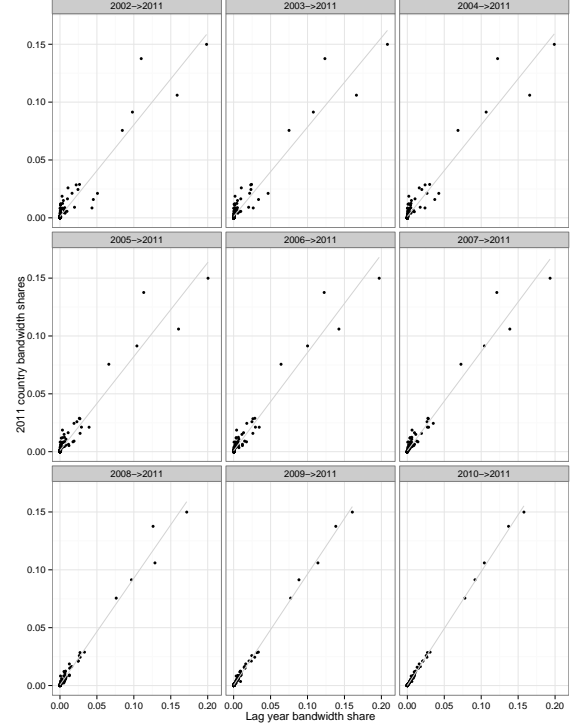


Figure 2. 2011 Country Bandwidth Proportions against 2002-2010 Country Proportions

Suppose instead of one distribution defining the overall distribution of bandwidth there are some distinct groupings of countries that share statistically similar approaches. As outlined above, we assume the outcomes of the micro-processes⁵ in each country group are additive with the observed bandwidth for a country being the sum of the outcome of the associated micro-processes. If this were to be

³This is a reasonable assumption if a quantum of bandwidth is assumed to be 1Mbps. By 2002 there are 931,420 quanta allocated among the 184 countries; enough to be extremely close to the stable share distribution predicted by the simple urn model.

⁴Using MLE methods, the bandwidth tail sizes ranged from 178 in 2003 to 22 in 2011. Power law p values ranged from 0.11 to 1.0 where numerically low p -values indicate *low* support for a power law. The high p -values supporting a power law occurred in those cases where the tail contained the fewest cases. This leads to low confidence in being unable to reject a power law when the tail size is very small. In all cases results of applying Vuong's (Vuong, 1989) log-likelihood statistic to the data favored the lognormal over the power law though similar caveats apply regarding the tail size.

⁵Here again we mean micro relative to the observed bandwidth amounts.

the case, the observed distribution would be a mixture of the distributions for each of the groups and, given the additivity assumption, it is reasonable to view the overall distribution as a mixture of Gaussian distributions. In other words, we posit that each country's bandwidth at a point in time results from summing the effects of various lower level policy, economic, political, and social processes. We further assume that the effects of each of these micro-processes are normally distributed. If the net effect of summing these processes is zero, we would have the basic urn model. If, on the other hand, the summative effect is considerably different from zero, we would expect to see this reflected in patterned deviations from the urn model for those countries. We term this a *mixture model*.

To examine this we used cross-validation as described in Shalizi (2013). In our case this involved randomly selecting half of the data points in the 2011 bandwidth distribution and iterating through groupings of size two to ten and selecting as K , the number of groups/parameters for which the log-likelihood is smallest, using an Expectation-Maximization algorithm (*normalmixEM*) provided in Benaglia, Chauveau, Hunter, and Young (2009)'s *mixtools R* package. This resulted in seven groups. Figure 3 shows the match between the theoretical cumulative distribution function (as identified in the training) and the empirical 2011 bandwidth distribution. While the fit, even in the tail, looks good, fitting 184 points with 21 parameters (7 lambdas, 7 means, and 7 standard deviations) is, of itself, not surprising. However, our goal here is to empirically identify micro-process groupings which can be used to understand changes in bandwidth shares over the period. This mixture produces the right skewed bandwidth distribution without resorting to power laws or lognormals.

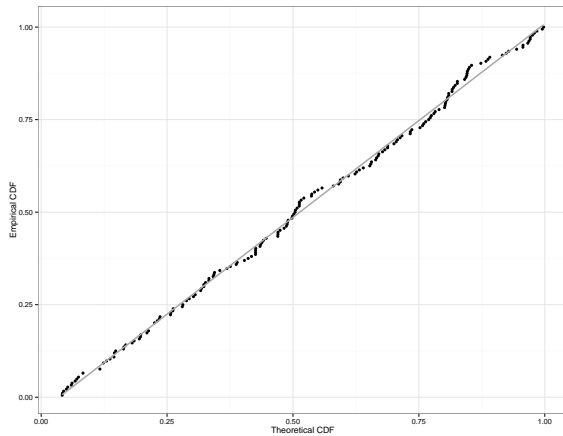


Figure 3. Theoretical (cross-validated) CDF v. 2011 Empirical CDF

Table 1 shows summary statistics for each of the seven empirically identified groups. In the table the *lambda* col-

umn contains the weights for each of the distributions (see Equation 1); they, of course, sum to 1.0. Given both that the basic urn model predicts stability over time and that it generally takes time for policy effects to be observed, we chose to then evaluate the effect of adding the Gaussian mixtures against the 2002 bandwidth data. $\Delta_{2011,2002}$ represents the arithmetic difference between a country's bandwidth share in 2011 and that in 2002. It will be positive where share increased over the 10 years and negative if it decreased. The $mean(\Delta_{2011,2002})$ and $var(\Delta_{2011,2002})$ are for each country group. The group percentile means (*Pctile Mean*) and number of countries (*Countries*) for each group are also reported.

Table 1
Summary Statistics for Country Groups, 2011

Group	λ	$mean(\Delta_{2011,2002})$	$var(\Delta_{2011,2002})$	$mean(Pctile)$	Cou
1	0.11	-0.0000003	0.0000000	6	2
2	0.16	0.0000022	0.0000000	20	2
3	0.16	0.0000252	0.0000000	37	3
4	0.22	0.0001801	0.0000000	56	3
5	0.12	0.0011358	0.0000002	72	2
6	0.21	0.0015709	0.0001274	88	3
7	0.03	-0.0178492	0.0011056	99	

Note: $\Delta_{2011,2002}$ refers to difference between 2011 bandwidth proportion and that in 2002. Zero values result from rounding.

Figure 4 provides a graphic visualization of the relationship between *Diff* and the 184 countries organized by group. The points on the plot are jittered to reduce over plotting and a local regression (LOESS) curve is drawn to highlight changes in the data. Recall that the simple urn model would predict that *Diff* would be 0. That is, the 2011 share would be the same as that in 2002. For countries in Groups 1 through 4 this is pretty much the case; for each of these groups $mean(\Delta_{2011,2002})$ is very close to 0. Groups 1 through 4 include 121 of the 184 countries. Groups 5 and 6 show an average gain in share and Group 7 a loss. Both Groups 6 and 7 have much higher variances than the other groups. Note from Table 1 that higher numbered groups contain higher bandwidth percentile countries.

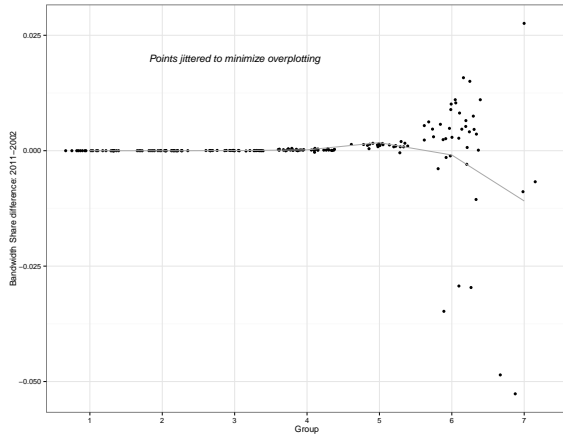


Figure 4. Country Bandwidth Share Differences ($\Delta_{2011,2002}$) organized by Group

An interpretation of this is that countries below about the 70th percentile have seen little change in bandwidth share. For example, countries in Africa, with the exception of Egypt, are in Groups 1 through 4 indicating that, on average, they have seen little share growth over the period. Nigeria, South Africa, and Algeria, three relatively strong African economies, all saw their 2011 shares virtually identical to those in 2002. Egypt, in Group 5, did show a slight share increase over the period.

Groups 5 and 6 (those above the 70th percentile but below the 99th) do show, on average, increased share. These include countries such as China (averaging double-digit growth over the period) and the Russian Federation (whose economy grew seven percent annually between 2000 and 2008) which, during the period, have made concerted efforts to connect more to the global economy and, importantly, had enjoyed significant resource increases permitting them to succeed. The largest share gaining country is Germany (the outlier in Group 7) which saw its economy rebounding from the costs of unification during this period (Dustmann, Fitzenberger, Schönberg, & Spitz-Oener, 2014).

These gaining countries also benefited by themselves being more economically attractive thus inducing other countries to connect to them thus increasing their bandwidth. This can be seen by looking at the global Internet from a network perspective. Russia's normalized eigenvector centrality grew from .01 in 2002 to .10 by 2011. China's went from .06 (2002) to .14 (2011) and Germany's jumped from .46 in 2002 to .94 in 2011. Over the same period, the normalized eigenvector centrality score for the United States (the biggest Group 7 loser) dropped from 1.0 in 2002 to .82 in 2011. Moreover, other research (Author, 2015) has shown that as countries get new bandwidth it tends to be via connections on the same continent. This favors European countries and reinforces the difficulty in seeing share change in Africa.

More than a decade after its economic reforms in the late 1970s and early 1980s, China started investing in Internet use and development eventually rising to the top of the world in 2008 in terms of the number of Internet users (Liang & Lu, 2010). Its first international line to the Internet was connected in 1994, and these early networks were dedicated mainly for academic exchange of information. As China quickly realized the importance of the Internet in economic development, the number of Internet users in the country increased exponentially from 2 million in 1998 to 100 million in 2005 to 650 million in 2014 (Internet World Stats, 2014). Its Internet bandwidth amount has increased from 20,228 Mbps in 2002 to 2,857,095 Mbps in 2011 for a CAGR of 73 percent over the period. This compares with 52 percent CAGR for the U.S. and 102 percent for Chile over the same period (TeleGeography, 2012). China subsequently enhanced its control and regulation over Internet infrastructure and censorship over political and social use of the Internet – a phenomenon widely known as the “Great Firewall of China” (Liang & Lu, 2010).

As a major economic hub in Europe, Germany has served as a main connecting point for fiber-optic networks in Europe and around the world. For example, in 2014 Finland announced that it is building a submarine cable at the bottom of the Baltic Sea connecting directly to Germany to increase available bandwidth between Finland and the rest of the European continent (Sverdlik, 2014). Domestically, Germany has fueled its continued growth of the Internet with fast broadband services in big cities and unmetered access products (Internet World Stats, 2014). To address the problem of slower Internet connections outside the big cities, Germany has recently adopted Digital Agenda 2014-2017 aimed at expanding fast broadband Internet both in rural and urban areas by 2018 (Deutsche Welle, 2014).

Group 6 countries showing a decrease, even as the group average increased, included Ireland, Denmark and South Korea. Interestingly these are all countries which had made connectivity a national priority in the late 20th and early 21st centuries. We may now be seeing a sort of regression effect where bandwidth shares of such previously high effort countries are declining a bit.

From the digital divide perspective, most interesting are the Group 7 countries. These are at the very top of the bandwidth distribution and, as a group, showed the only average drop in share. The richest countries, with the exception of Germany, lost rather than gained share. That loss, however, was balanced by average gains in Groups 5 and 6.

Average shares for the bandwidth poorest – Groups 1 through 4 – remained unchanged and consistent with the basic urn model. The share of total international bandwidth held by the top one percent of countries decreased from 36 percent in 2002 to 29 percent in 2011. The top five percent share dropped from 84 percent to 69 percent, and the top 10

percent dropped slightly from 84 percent to 83 percent. The share of the top half of countries remained fairly constant moving from 99.9 percent to 99.7 percent.

Discussion and Conclusion

We adopted a network perspective with country as the unit of analysis from which to examine the evolution of country level international Internet bandwidth over the 2002-2011 interval. The distributions were consistently asymmetric with a very heavy right tail. In general, all countries have gotten increasingly more bandwidth while maintaining shares of total bandwidth in a manner compatible with our preferential attachment based *mixture model*. Here *mixture model* refers to viewing country bandwidth as resulting from a mixture of Gaussian distributed political, economic, policy, and social micro-processes (Equation 1) with parameters estimated using the cross-validation approach described in the previous section. When the net additive effect of these micro-processes is zero, we have the simple urn model. This appears to be descriptive of countries below about the 70th percentile in bandwidth share. For these countries, as expected from the simple urn model, bandwidth shares remained fairly constant over the 2002-2011 period. However, countries above the 70th percentile do show considerable deviation as would be expected from non-zero net effects from the micro-processes. As seen, this *mixture model* is also capable of producing the observed right skewed bandwidth distribution. Finally, it is important to note that our model captures the notion of bandwidth distributions being generated via a preferential attachment mechanism without also implying that bandwidth be power law distributed. The logic flow behind our mixture model is shown in the Appendix.

More generally, our most robust empirical result is that the bandwidth share distributions, and thus bandwidth differences, exhibit considerable path dependence. For most countries, global Internet bandwidth shares have been remarkably similar over the period. While it is true that the bandwidth richest countries have increased their bandwidth over the period, the bandwidth poorer ones have experienced increases as well. Though the raw bandwidth gap between the richest countries and the poorest ones has increased, the proportion of total international bandwidth held by the five bandwidth richest countries has actually decreased and that of the next tier, 70th to 97th percentiles, has shown modest increases. In other words, share gaps among the 63 bandwidth richest countries have narrowed slightly, and shares among the 121 bandwidth poorest countries have remained constant over the period.

Moreover, while beyond the scope of the research we report here, we see some evidence that increased bandwidth is not simply the result of increased domestic demand but rather increases in bandwidth in many cases precede increased demand. That is, increasing international bandwidth is an affor-

dance that may help to bring about increased Internet usage at future time points. Build it and they will, with increased probability, compute. From the standpoint of our model and given the relative rarity of a new country entering into the global Internet a key to a bandwidth poor country overcoming the inertia of its historical position almost certainly would require a concerted effort of the sort made by South Korea to not only increase bandwidth but to increase its share of global bandwidth beyond that resulting from preferential attachment. On the other hand, as the relative values of information and knowledge continue to grow, there is risk that today's bandwidth share inequalities will harden into dramatic information inequalities separating the bandwidth rich and bandwidth poor countries.

From a policy perspective, our analysis underscores the importance of carefully specifying the gap being referred to when speaking of the global digital divide. While this is in part a normative question, it does appear that strategies aimed at expanding the lower half of the distribution's share of bandwidth will require different measurement metrics and approaches than might those targeting growing absolute gaps at the top end of the distribution given the heavy inertial effect of the 2002 distribution. In particular, countries in Groups 1 through 4 are precisely the countries that tend to lack the economic and/or political capacity to implement successful policies directed at increasing bandwidth beyond that expected by preferential attachment. Moreover, when viewed as a network, it is clear that what one country does must be considered within the context of what other countries are doing. Covariance is a feature of networks. As a consequence, policy impacts, especially for the bandwidth smaller ones, are somewhat constrained by the policies of others in the network. For example, Africa countries have remained among the least connected and it will be increasingly difficult for them to catch up any time soon if the preferential attachment component of our model continues to dominate within the lower ranked country groupings. That this has important consequences for low bandwidth countries' political, economic, and societal developments has been well documented (Oyelaran-Oyeyinka & Adeya, 2004; Roycroft & Anantho, 2003).

As is case with any empirical research, our study is not without limitations. First, we relied mainly on one dataset – international Internet bandwidth. While, as discussed above, this is a good measure of international Internet capacity, this type of research could be enriched by taking into account other relevant datasets such as mobile communication networks. Second, our country-level analysis did not consider country differences in such things as population. For example, while per-capita analyses would be interesting, our objective in this paper was to examine the global structure of the Internet. Third, we focused on structural and distributional elements as they are key to understanding global dig-

ital divide and also there has been insufficient academic attention to this topic. However, analyzing both structural aspects and more traditional country-specific political and economic characteristics would be helpful. In particular, focused country-level case studies designed to ascertain whether our country groupings make sense under fine grained examination would be helpful. Further, examining neighborhood structures in regions or continents may provide useful information. Finally, bandwidth research such as ours would benefit from being able to make meaningful comparisons of total (as opposed to proportion shares) bandwidth across time. Hilbert and López (2012)'s work is promising here though much remains to be done before we have a credible way of adjusting bandwidth to constant levels.

Information technology has reshaped economies and societies around the world. Understanding the distribution and evolution of these changes is important to identifying potential problems and opportunities and then developing appropriate policies. Our results help in discriminating competing theoretical perspectives on the global digital divide. They also sharpen policy discussions related to gaps between bandwidth rich and bandwidth poor countries. This study makes theoretical and methodological contributions relevant to scholars and policymakers working in the area of international communication networks and the digital divide. In particular, our development of a preferential attachment based mixture model of bandwidth allocation provides a more nuanced picture of the digital gaps across the globe and their implications for other aspects of society. The approaches used in this study may be applied to studying digital divide within a country or society. Future research may also consider this interdisciplinary analysis perspective to better understand specific issues in international communication and digital media-based communication.

Appendix Urn-Based Mixture Model Logic

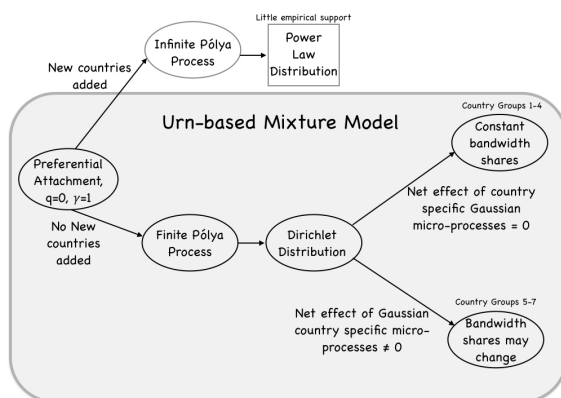


Figure A1. Urn-Based Mixture Model

References

- Andrés, L., Cuberes, D., Diouf, M., & Serebrisky, T. (2010). The diffusion of the internet: A cross-country analysis. *Telecommunications Policy*, 34(5), 323–340.
- Arrow, K. J. (2000). Increasing returns: historiographic issues and path dependence. *European Journal of the History of Economic Thought*, 7(2), 171–180.
- Arthur, W. B. (1994). *Increasing returns and path dependence in the economy*. University of Michigan Press.
- Asia Pacific Economic Cooperation. (2002). New economy in APEC: Innovations, digital divide, and policy. *Asia-Pacific Economic Cooperation (APEC) Secretariat*, Singapore.
- Barabási, A.-L. (2003). *Linked: How everything is connected to everything else and what it means*. Penguin Group New York.
- Barnett, G. A. (2001). A longitudinal analysis of the international telecommunication network, 1978–1996. *American Behavioral Scientist*, 44(10), 1638–1655.
- Barnett, G. A., & Park, H. W. (2005). The structure of international internet hyperlinks and bilateral bandwidth. *Annals of Telecommunications*, 60(9), 1110–1127.
- Barnett, G. A., & Park, H. W. (2014). Examining the international internet using multiple measures: new methods for measuring the communication base of globalized cyberspace. *Quality & Quantity*, 48(1), 563–575.
- Barnett, G. A., Ruiz, J. B., Hammond, J. R., & Xin, Z. (2013). An examination of the relationship between international telecommunication networks, terrorism and global news coverage. *Social Network Analysis and Mining*, 3(3), 721–747.
- Benaglia, T., Chauveau, D., Hunter, D. R., & Young, D. (2009). mixtools: An R package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6), 1–29. Retrieved from <http://www.jstatsoft.org/v32/i06/>
- Benkler, Y. (2006). *The wealth of networks: how social production transforms markets and freedom*. New Haven: Yale University Press.
- Bonfadelli, H. (2002). The internet and knowledge gaps a theoretical and empirical investigation. *European Journal of communication*, 17(1), 65–84.
- Castells, M. (2011). *The rise of the network society: The information age: Economy, society, and culture* (Vol. 1). Wiley-Blackwell.
- Chinn, M. D., & Fairlie, R. W. (2006). The determinants of the global digital divide: a cross-country analysis of computer and internet penetration. *Oxford Economic Papers*.
- Chinn, M. D., & Fairlie, R. W. (2010). Ict use in the developing world: an analysis of differences in computer and internet penetration. *Review of International Economics*, 18(1), 153–167.
- Chung, F., Handjani, S., & Jungreis, D. (2003). Generalizations of Polya's urn problem. *Annals of combinatorics*, 7(2), 141–153.
- Clauset, A., Shalizi, C., & Newman, M. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661–703.
- Dasgupta, S., Lall, S., & Wheeler, D. (2005). Policy reform, economic growth and the digital divide. *Oxford Development Studies*, 33(2), 229–243.

- Deutsche Welle. (2014). German cabinet adopts "digital agenda" for expanded broadband Internet. Retrieved from <http://www.dw.de/german-cabinet-adopts-digital-agenda-for-expanded-broadband-internet/a-17867699>
- Dustmann, C., Fitzenberger, B., Schönberg, U., & Spitz-Oener, A. (2014). From sick man of Europe to economic superstar: Germany's resurgent economy. *The Journal of Economic Perspectives*, 28(1), 167–188.
- Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 1429–1451.
- Eggenberger, F., & Pólya, G. (1923). Über die statistik verketteter vorgänge. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 3(4), 279–289.
- Fink, C., & Kenny, C. J. (2003). W (h)ither the digital divide? *info*, 5(6), 15–24.
- Frank, S. A. (2014). How to read probability distributions as statements about process. Available at SSRN 2496734.
- Galor, O. (1996). Convergence? inferences from theoretical models. *The Economic Journal*, 1056–1069.
- Giddens, A. (1990). The consequences of modernity. *Cambridge: Polity*.
- Guillén, M. F., & Suárez, S. L. (2005). Explaining the global digital divide: Economic, political and sociological drivers of cross-national internet use. *Social Forces*, 84(2), 681–708.
- Hafner, K., & Lyon, M. (1999). *Where wizards stay up late: The origins of the Internet*. Simon and Schuster.
- Hassani, S. N. (2006). Locating digital divides at home, work, and everywhere else. *Poetics*, 34(4), 250–272.
- Heeks, R., & Kenny, C. J. (2001). Is the internet a technology of convergence or divergence? *World Bank, Washington, DC Processed*.
- Hilbert, M. (2014). Technological information inequality as an incessantly moving target: The redistribution of information and communication capacities between 1986 and 2010. *Journal of the Association for Information Science and Technology*, 65(4), 821–835.
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60–65.
- Hilbert, M., & López, P. (2012). How to measure the world's technological capacity to communicate, store and compute information? part ii: Measurement unit and conclusions. *International Journal of Communication*, 6, 936–955.
- Howard, P. N. (2010). *The digital origins of dictatorship and democracy: Information technology and political Islam*. Oxford Univ Pr on Demand.
- Howard, P. N., & Hussain, M. M. (2011). The role of digital media. *Journal of Democracy*, 22(3), 35–48.
- Hyun, K. (2012). Americanization of web-based political communication? a comparative analysis of political blogospheres in the United States, the United Kingdom, and Germany. *Journalism & Mass Communication Quarterly*, 89(3), 397–413.
- International Telecommunications Union. (2012). *ITU measuring the information society 2012 report*.
- Internet World Stats. (2014). Internet world stats. Retrieved from <http://www.internetworldstats.com>
- Ioannides, Y. M. (2013). *From neighborhoods to nations: The economics of social interactions*. Princeton University Press.
- James, J. (2007). From origins to implications: Key aspects in the debate over the digital divide. *Journal of Information Technology*, 22(3), 284–295.
- Jovanovic, B., & Rob, R. (1989). The growth and diffusion of knowledge. *The Review of Economic Studies*, 56(4), 569–582.
- Kiiski, S., & Pohjola, M. (2002). Cross-country diffusion of the internet. *Information Economics and Policy*, 14(2), 297–310.
- Lee, C., & Chan-Olmsted, S. M. (2004). Competitive advantage of broadband internet: a comparative study between South Korea and the United States. *Telecommunications Policy*, 28(9), 649–677.
- Lee, S., Monge, P., Bar, F., & Matei, S. (2007). The emergence of clusters in the global telecommunications network. *Journal of Communication*, 57(3), 415–434.
- Liang, B., & Lu, H. (2010). Internet development, censorship, and cyber crimes in China. *Journal of Contemporary Criminal Justice*, 26(1), 103–120.
- Mahmoud, H. (2008). *Pólya urn models*. CRC press.
- Manyika, J., & Roxburgh, C. (2011, October). *The great transformer: The impact of the Internet on economic growth and prosperity* (Tech. Rep.). McKinsey Global Institute.
- Monge, P., & Contractor, N. (2003). *Theories of communication networks*. Oxford University Press, USA.
- Mossberger, K., Tolbert, C. J., & Hamilton, A. (2012). Broadband adoption| measuring digital citizenship: Mobile access and broadband. *International Journal of Communication*, 6, 37.
- Newman, M., Barabási, A.-L., & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton University Press.
- Norris, P. (2003). *Digital divide: Civic engagement, information poverty, and the internet worldwide* (Vol. 40). Taylor & Francis.
- Oyeleran-Oyeyinka, B., & Adeya, C. (2004). Internet access in Africa: empirical evidence from Kenya and Nigeria. *Telecommunications and Informatics*, 21(1), 67–81.
- Park, H. W., Barnett, G. A., & Chung, C. J. (2011). Structural changes in the 2003–2009 global hyperlink network. *Global networks*, 11(4), 522–542.
- Pejovic, V., Johnson, D. L., Zhelva, M., Belding, E., Parks, L., & Van Stam, G. (2012). The bandwidth divide: Obstacles to efficient broadband adoption in rural sub-Saharan Africa. *International Journal of Communication*, 6, 2467–2491.
- Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American political science review*, 251–267.
- Piketty, T. (2014). *Capital in the twenty-first century*. Harvard University Press.
- Quibria, M. G., Ahmed, S. N., Tschang, T., & Reyes-Macasaquit, M.-L. (2003). Digital divide: Determinants and policies with special reference to Asia. *Journal of Asian Economics*, 13(6), 811–825.
- Roycroft, T., & Anantho, S. (2003). Internet subscription in Africa: policy for a dual digital divide. *Telecommunications Policy*, 27(1), 61–74.
- Shalizi, C. (2013). *Advanced data analysis from an elementary point of view*. (2013 draft)

- Standage, T. (1998). *The Victorian internet: the remarkable story of the telegraph and the nineteenth century's online pioneers*. Weidenfeld & Nicolson London.
- Sverdlik, Y. (2014). Finland building direct submarine cable to Germany. Retrieved from <http://www.datacenterknowledge.com/archives/2014/12/04/submarine-cable-to-germany-to-boost-finland-internet-bandwidth/>
- TeleGeography. (2012). *Global Internet geography* (Tech. Rep.). PriMetrica, Inc.
- Van Dijk, J. (2005). *The deepening divide: Inequality in the information society*. Sage Publications, Incorporated.
- Vinciguerra, S., Frenken, K., & Valente, M. (2010). The geography of internet infrastructure: an evolutionary simulation approach based on preferential attachment. *Urban Studies*, 47(9), 1969–1984.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307–333.
- Wong, P.-K. (2002). Ict production and diffusion in asia digital dividends or digital divide? *Information Economics and Policy*, 14(2), 167–187.
- Wunsch-Vincent, S., & Vickery, G. (2010). *The evolution of news and the internet*. OECD.
- Yu, L. (2006). Understanding information inequality: Making sense of the literature of the information and digital divides. *Journal of Librarianship and Information Science*, 38(4), 229–252.
- Zenou, Y., & Jackson, M. (2013). Economic analyses of social networks.